DOCUMENT RESUME

ED 173 331 SP 014 407

AUTHOR Nickel, K. N.
TITLE Experimentation, Extrapolation, Exaggeration--Thy Name is Research.
PUB DATE [79]
NOTE 9p.

EDRS PRICE MF01/PC01 Plus Postage.
DESCRIPTORS *Bilingual Education; Case Studies; Control Groups; Data Collection; *Educational Research; Language Dominance; *Research Methodology; *Research Problems; Sampling; Socioeconomic Status; Statistical Analysis; Teacher Characteristics
IDENTIFIERS *American Institutes for Research; Bilingual Education Act 1968

ABSTRACT

The failure of educational research to fully allow for limitations in methodology when drawing conclusions is faulted. In addition to this, a failure on the part of the media--educational and other--to fully publicize these limitations is also criticized. Methodological shortcomings of the 1977-78 American Institutes of Research (AIR) study of Title VII bilingual education are used as an example of faulty research methodology and subsequent distortion of the results of that research by media. Guidelines for evaluating the worth of educational research in reference to six possible limitations are presented. (MJB)

****
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
****

ED173331

U S DEPARTMENT OF HEALTH, EDUCATION & WELFARE
NATIONAL INSTITUTE OF EDUCATION

THIS DOCUMENT HAS BEEN REPRODUCED EXACTLY AS RECEIVED FROM THE PERSON OR ORGANIZATION ORIGINATING IT. POINTS OF VIEW OR OPINIONS STATED DO NOT NECESSARILY REPRESENT OFFICIAL NATIONAL INSTITUTE OF EDUCATION POSITION OR POLICY

"PERMISSION TO REPRODUCE THIS MATERIAL HAS BEEN GRANTED BY

K. N. Nickel

TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)."

EXPERIMENTATION, EXTRAPOLATION, EXAGGERATION---THY NAME IS RESEARCH

By K. N. Nickel

The perceptive scholar who uttered the immortal words, "Anything can be proved by statistics," not only was concerned with research in many fields--psychology, science, education and others--and with both applied and basic research, this scholar also knew "Madison Avenue" advertising types and the current political declaimers and exclaimers. Statistics from various research studies are being used and misused at an astounding rate today to "prove" product reliability, certain hypotheses, and political greatness--more than ever before. It appears that each successive newscast or publication includes glaring examples of "proving anything" by extrapolation and/or exaggeration of data from experimentation in research.

Statistics and tentative conclusions from preliminary reports or interim studies are trumpeted widely without regard to the tentativeness, the limitations of the study (and every study has some), the composition and/or personality of the population sample, the validity and/or reliability of the procedures and controls in the study (or lack thereof), and the narrowness of applicability of the findings. It seems that few, if any, of the trumpeters take time to examine the research design and the organization of the study before blindly applying the tentative conclusions to a broad population outside the realm of the study. Tentative conclusions as well as final reports have been exaggerated far beyond the scope of many studies and have been magnified to distortion in many cases.

It is true that research with medicines, plants, chemicals, animals, machines, and people is necessary for progress to continue. However, it is equally true that distorted statements because of incorrect or inaccurate research or, more frequently, because of fallacious and unprecise extrapolation from research data can inhibit progress or even reverse it. White[1] in

SP 011 407

his review of Pseudoscience and Mental Ability emphasized this point:

> That search, in its early as well as contemporary phases, used distorted or invented data to support social and racial prejudice.

Fortunately, deliberate distortion or invention of false data is rare. However, exaggerated statements of research results can be as damaging as false reports.

Researchers know that data from any research study apply only to that study under the conditions that existed at the time the data were collected. And when the research is done with human beings, it is much more difficult to control and/or stipulate certain conditions because of the many differing influences from past experiences and the immediate environment that impinge on the persons in the study. Unfortunately, those who publicize the results of a particular research study seldom know (or seem to care about) the limitations with regard to the sample population or the conditions of the study, the influences that affected the subjects in the sample population, or the qualifying factors which were instrumental in determining conclusions which were reached. Incomplete or inaccurate statements about a study are broadcast and re-broadcast widely and are soon quoted as fact. Such is the status of much of the information about research studies today.

The current situation with regard to a recent study is a good example of today's treatment of research. The headline on an article in the September, 1978, issue of Phi Delta Kappan, "Lower Achievement Found in Bilingual Programs,"[2] matches other headlines on this study. "A Storm Brews Over Bilingual Teaching"[3] and "Bilingualism: where to draw the line"[4] are other examples. The Kappan article as with the others repeats as fact some inaccurate data. Newspapers have carried numerous columns which are similar, radio and television news persons have duly reported the story,

state legislative committees and Congressional committees have deliberated over these statements, and magazines have included lengthy articles on this study. They all report research findings showing a lack of effectiveness of bilingual education based on the American Institutes of Research (AIR) Report which was financed by the US Office of Education. This study was to evaluate bilingual education programs with an experimental group of students in Title VII programs and a control group of students in regular classes. The extrapolation and exaggeration began with an Interim Report from AIR in April, 1977, and has continued and grown with publication widespread. The AIR Report is cited as "proving that bilingual education is a failure." A review of some of the details of this study is necessary if the results are to be used as criteria to judge the effectiveness of bilingual education.

The AIR population sample included 5800 students in Title VII Spanish/English bilingual projects and 2400 non-Title VII comparable students in terms of (a) ethnicity, (b) socioeconomic status, and (c) grade level.[5] However, there is some question with regard to the comparableness of the two groups as the Interim Report included the following statement:

> The non-Title VII students were roughly comparable to the Title VII students, with the one exception that more non-Title VII students than Title VII students were classified by their teachers as either English-dominant or bilingual English-dominant--96 per cent as compared to 75 per cent.[6]

The Final Report Executive Summary issued in March, 1978, did not include this very important difference although the two reports were almost word-for-word the same otherwise. The following statement replaced the statement given earlier from the Interim Report:

> The non-Title VII students were found to be basically comparable to the Title VII students and, in the analysis, standard statistical procedures were used to adjust for differences between the two groups.[7]

It is noteworthy that a rather detailed search of the several volumes of this report failed to find any mention of the differences in English ability reported in the Interim Report and could discover no discussion of statistical treatment of these differences.

The AIR Report does state that "Fifty per cent of the Title VII teachers and sixty-six per cent of their aides said that they were currently proficient in both Spanish and English."[8] It appears from this statement that fifty per cent of the teachers were not considered proficient in Spanish and English. No mention could be found in the study of this serious limitation since adequate bilingual education requires that each teacher be bilingual.

The Evaluation Study--Executive Summary (Final Report) issued in March, 1978, describes the final results of the study--"results describing the effectiveness of the ESEA Title VII Spanish/English bilingual education program as of the 1975-76 academic year."[9] A footnote on this page states: "The final results of this evaluation are similar to interim results of the same evaluation released in April 1977."[10] It seems logical to conclude that the March, 1978, report did not include the "follow-up" testing in late fall 1976 but only included data from "The testing sessions . . . . . in the late fall and late spring of the 1975-6 school year."[11]

What do we find when we look at the sample population, the research design, and the limiting factors in the AIR study. We find students in bilingual education classes with inadequate teachers (at least 50 per cent) compared to a control group of students who were more competent in English. The students in the bilingual education classes were expected to catch or surpass the control group in English ability in a few months--"from late fall to late spring." When these students did not perform as well as or better than the control group, the headlines condemned all bilingual educa-

tion programs as failures. Gray[12] commented on this unfairness in the AIR study:

> Serious limitations in the study are attributed to the fact that the average period between the pre- and post-testing is five months or less. It seems unrealistic to expect either the rate of improvement or the actual improvement to be evident over such a short time span.

It appears that there are several serious limiting factors in the AIR study which were basic to the research design. However, little can be found in the statistical analysis to correct for these limitations. It could be assumed that the researchers felt that the differences in English ability, the less-than-adequate teachers, and the very short period of time were not as important as ethnicity, socioeconomic status, and grade level. However, most observers would protest the results vigorously if such unequal competition as this unfair comparison were used in other endeavors and the charges of bias, unfairness or worse would be heard. Yet, the results of the AIR study have been widely publicized, accepted and cited as "proving" the ineffectiveness of bilingual education.

Hopefully, those who receive, publicize and comment on research studies will attempt to find out more about the limitations, the design structure, and the conditions which affect and limit applicability of studies which come to their attention before making all-inclusive, far-reaching, exaggerated statements about such research. Until this is done, we will be subject to reports and comments that are incorrect and, perhaps, even dangerous to life and limb and particularly progress. Currently, it can only be said: Experimentation, Extrapolation, Exaggeration--Thy Name Is Research!

It would indeed be helpful to utilize carefully the criteria which have been presented in a recent publication for evaluating research.

These criteria were used to evaluate research studies in bilingual education and, incidentally, were instrumental in the conclusion that the AIR study was not sound research.

Research was considered inappropriate for discussions of the effectiveness of bilingual education on student performance if it demonstrated one or more of the following methodological weaknesses:

1. NO CONTROL FOR SOCIOECONOMIC STATUS
   In many research studies, students were compared without concern for disparities in their social and economic background. Frequently the performance of middle class monolingual children was compared with the performance of lower class bilingual children, to the obvious disadvantage of the lower class bilingual child.....

2. INADEQUATE SAMPLE SIZE/IMPROPER SAMPLING TECHNIQUES/EXCESSIVE ATTRITION RATE
   Any research study with a sample size of less than thirty students was considered inadequate. An inadequate sample made generalization to a larger population impossible. In some research studies and project evaluations, non-random selection of a sample population invalidated the findings presented. Often biased, "anecdotal" information was invalidly presented as representative of student performance.
   In addition, in some studies the number of students remaining in the posttest sample was less than 40% of the original sample.....

3. NO BASELINE OR COMPARISON DATA/NO CONTROL GROUP/NON-RELEVANT COMPARISONS
   Most often project evaluations attempted to demonstrate program effectiveness through statistical analysis of gain scores..... However, without a comparison or control group, or in the absence of some baseline measure of student gains prior to inception of the program, such statistical analyses are meaningless to discussions of program effectiveness.....

4. NO CONTROL FOR INITIAL LANGUAGE DOMINANCE
   In many research and project evaluations, ethnic identity was equated with the language dominance of an individual student.... No attempt was made to ascertain the degree to which the child of a particular ethnic background was actually fluent in the language of that ethnic group.
   In other instances, an individual student's language ability was assessed by a teacher, often a monolingual teacher unfamiliar with the other language being assessed. Such teacher assessments have been shown to be highly unreliable means of determining language dominance.....

5. SIGNIFICANT DIFFERENCES IN TEACHER QUALIFICATIONS OR CHARACTERISTICS/OTHER CONFOUNDING VARIABLES
   In some research studies, a group of teachers were shown to have significantly better training, education and experience than teach-

ers of a comparison group. In other instances, certain programs had clearly superior teaching materials, classroom facilities, etc., than did the comparison program. In such cases, the effect of the differences in teacher qualifications, teaching materials, etc., could seriously bias the outcome of the programs being compared.....

6. INSUFFICIENT STATISTICAL INFORMATION/IMPROPER STATISTICAL APPLICATIONS

Frequently, project evaluations did not provide sufficient statistical information from which valid conclusions as to program effectiveness could be drawn.....

In some research and project evaluations, improper experimental designs were employed, or statistical procedures were improperly applied. In some instances, experimental designs were employed which called for the random assignment of students to treatment groups when, in fact, the students being studied were not randomly assigned to the treatment conditions. In other instances, analysis of covariance was misused. Theoretically, this technique adjusts for initial differences in the treatment and comparison groups. If these initial differences are large, however, analysis of covariance is ineffective in controlling for the differences involved.....[13]

These authors considered that the AIR Study did not meet criteria Nos. 4, 5 and 6 and, thus, should not be cited as effective research in the area of bilingual education.[14] These criteria with some accompanying explanation have been included in the hope that more critical examination will be made of research reports before the results are "trumpeted" throughout the land and before exaggerations become established fact.

REFERENCES

1. Edward M. White, "Myths About Measurement," The Chronicle of Higher Education, July 17, 1978, page 13.

2. "Lower Achievement Found In Bilingual Programs," Phi Delta Kappan, September, 1978, Vol. 60, No. 1, page 6.

3. "A Storm Brews Over Bilingual Teaching," U.S. News & World Report, March 6, 1978, page 58.

4. The Christian Science Monitor, May 22, 1978, page 15.

5. Evaluation Study, Executive Summary. Final Report, March 1978, page 3.

6. Evaluation Study, Executive Summary. Interim Report, April 1977, page 3.

7. Final Report, page 3.

8. Final Report, page 4.

9. Final Report, page 1.

10. Final Report, page 1.

11. Final Report, page 2.

12. Tracy C. Gray, "The AIR Report--Another View," Forum, The National Clearinghouse for Bilingual Education, Rosslyn, VA, Vol. 1, Issue 7, September, 1978, page 2.

13. L. T. Zappert & B. R. Cruz, Bilingual Education: An Appraisal of Empirical Research. Berkeley, CA: The Berkeley Unified School District, 1977, pages 4-6.

14. Ibid., page 42.